IDENTIFYING SOCIOECONOMIC PATERNS IN URBAN AREAS TROUGH FUZZY MODELING AND OBJECT ORIENTED IMAGE CLASSIFICATION

Alexandre G. Evsukoff, Adriano Vasconcelos, Alexandre P. de Carvalho and Carlos Beisl

COPPE/Federal University of Rio de Janeiro, Brazil
Rio de Janeiro, Brazil

1. INTRODUCTION

Population and incoming surveys provide a detailed description of urban areas revealing high quality information for planning and decision making. Traditional censuses are, however, very costly and are thus released only in very long time periods (average 10 years in Brazil). The integration of geo-referenced censitary data into Geographic Information Systems (GIS) allows the analysis of the spatial dependencies of socioeconomic data, but it is still dependent on the expensive surveys.

Remote sensing imagery is a powerful tool for urban monitoring allowing the identification and detection of land use and landscape changes. High resolution images within modern classification algorithms allow the identification of the main urban features at a reasonable cost. This work presents a modeling approach to identify socioeconomic patterns in urban areas trough remote sensing images. The model relates image attributes to socioeconomic censitary data in a fuzzy rube based classification algorithm.

The preliminary results of the proposed approach have been obtained using SPOT-5 images and income information provided by the Brazilian Geographic and Statistic Institute (IBGE). The fuzzy rule base is obtained by a recently proposed fuzzy rule based classifier induction algorithm [1]. The application under study is the monitoring of the development impact of the implantation of the Rio de Janeiro State Petrochemical Complex (COMPERJ), which was announced in 2006 and should start operation in 2010. This work presents preliminary results of a long-term project for approximately identify the urban changes within the COMPERJ area trough remote sensing images.

2. PROJECT AND DATA DESCRIPTION

The COMPERJ complex is under construction in the Rio de Janeiro state by the Brazilian Petroleum Company (Petrobras) and will be a huge industrial plant aiming to process the crude oil produced at the Campos Basin, which correspond to 80% of current oil production in Brazil. The COMPERJ complex is expected to create about 200 thousands of direct and indirect working positions influencing 11 neighborhood districts in the Rio de Janeiro state.

Such a huge and ambitious enterprise should be followed by a careful urban development planning and monitoring. The method described in this work is a part of a long term project for urban development monitoring. The main objective of the project is to build a monitoring model for the entire influence area of the COMPERJ complex (11 districts) based on high resolution remote sensing imagery and as much as possible information gathered from several sources.

This paper describes the preliminary results obtained in a pilot application for the São Gonçalo district, using SPOT-5 data with 10 and 2.5 meters of spatial resolutions. The images taken in 2004 and 2005 were related to income and population data collected at the 2002 census released by IBGE. The São Gonçalo district is divided in 1220 censitary sectors, each of which containing about 200 domiciles, and whose cartographic information is also available from IBGE.

3. THE MODELING WORKFLOW

The modeling workflow is shown in Fig. 1. The geo-referenced data of São Gonçalo censitary sectors were loaded into the GIS software and the GIS database was exported to the image processing software to guide image segmentation. In regions of high population density the censitary sector may become very small to contain enough image information, such that the censitary sectors had to be clustered into segments of sufficient area. A total of 250 segments were obtained from the 1220 censitary sectors. The image segmentation was computed with Definiens software that executes a multi-resolution segmentation algorithm [2] such that several levels of segmentations with different granularities can be used for image classification. The Definiens software is also used to compute image attributes that could be related to socioeconomic patterns. The detailed description of image attributes will be presented in the full version of this paper. In the preliminary results, a total of 17 image attributes computed for all 250 segments were saved into a dataset. The image attributes dataset is joined with the IBGE income data and were divided into 2 classes; the low socioeconomic potential class, represented by the

33% lowest income values and the medium-high socioeconomic potential, contained the remaining data. Other partitions were also tested and will be presented in the final version of the paper. The symbolic fuzzy modeling algorithm [1] was used to generate a set of fuzzy rules that relates image attribute to income classes. The desired classification and the results of the fuzzy classifier were exported to GIS allowing the visualization of the uncertainty associated to the classification. The fuzzy rules were also implemented into Definiens software and used to classify an image segmented at a level of smaller granularity than the one used for training. The preliminary results are presented next.

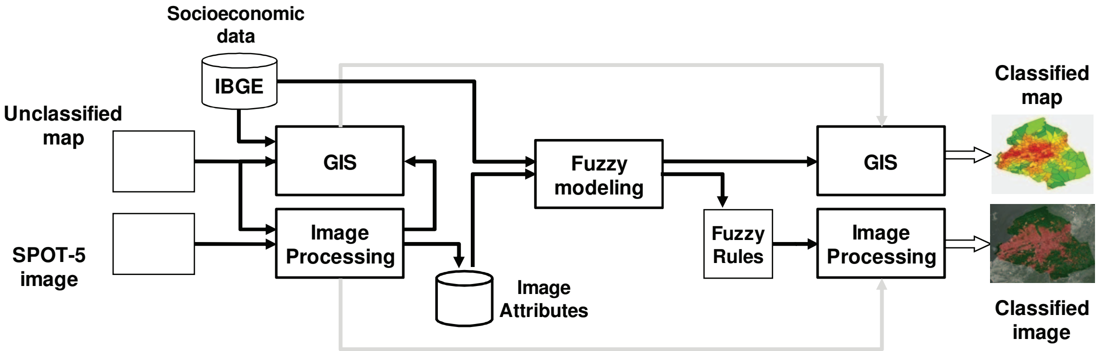
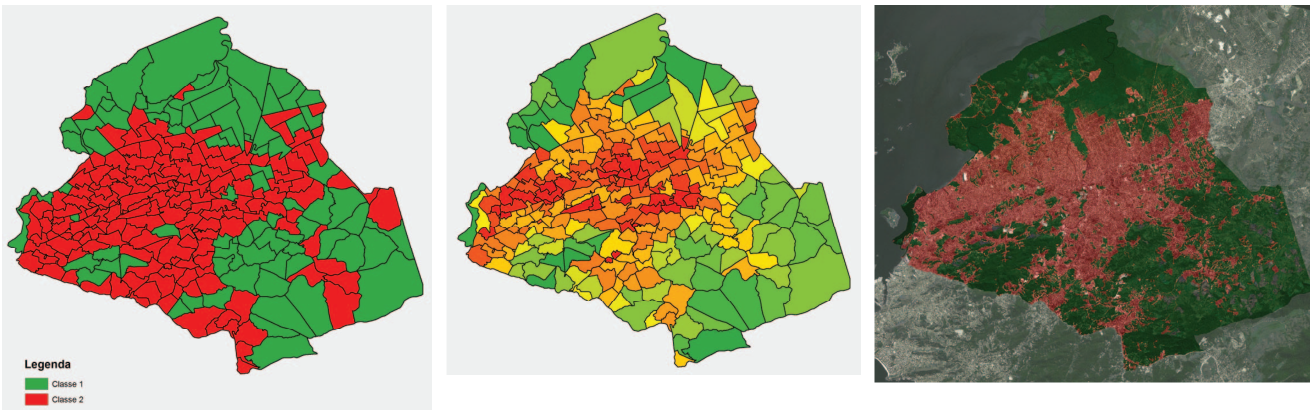


Fig. 1: The modeling workflow.

4. PRELIMINARY RESULTS

The symbolic fuzzy model for the data described in this abstract has identified 4 rules using only 5 image attributes. In the full version of the paper, the fuzzy rules will be described in detail. The fuzzy classifier has obtained 79% of agreement but, due to the imbalance of the dataset, the kappa statistic of only 0.5 indicates a moderate result. Nevertheless, the visualization of the results agrees to the expectation of the experts. The preliminary results can be visualized in Fig. 2. In Fig. 2(a) the GIS model of the income classification in two classes: the low income (green), which represents 33% of the segments and medium-high income (red) that correspond to 66% of the segments. The fuzzy classification results are shown in Fig. 2(b) where the membership values are represented by color grades, indicating uncertain classification. The image classification computed by the Definiens software is shown in Fig. 2(c), where a good visual correspondence with the desired classification is achieved. The full paper will present further results of the other classes' partition that have been studied.



(a) (b) (c)

Fig. 2: Preliminary results. (a) Desired classification of the 250 segments divided in medium-high (red) and low (green) income; (b) Fuzzy model results, showing partial membership as color shades; (c) Image classification result.

5. REFERENCES

[1] A.G. Evsukoff, S. Galichet, B. S. L. P. de Lima and N. F. F. Ebecken, "Design of interpretable fuzzy rule-based classifiers using spectral analysis with structure and parameters optimization". Fuzzy Sets and Systems (2008), doi: 10.1016/j.fss.2008.08.010.

[2] U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, M. Heynen, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information" ISPRS Journal of Photogrammetry & Remote Sensing 58, pp. 239– 258, 2004.